Precision in Weighing: A Comparison of Scales Found in Physician Offices, Fitness Centers, and Weight Loss Centers

RISA J. STEIN, PHD^a
C. KEITH HADDOCK, PHD^b
WALKER S.C. POSTON, PHD,
MPH^b
DANA CATANESE^c
JOHN A. SPERTUS, MD^b

SYNOPSIS

Objective. Weight assessment is a critical aid in patient care. It is particularly important in monitoring progression of pregnancies, heart failure status, and when adjusting medications. Although weight is generally determined using a scale, few studies have evaluated the precision of non-household scales. The objective of this study was to assess scale precision across a variety of settings.

Methods. An evaluation of scales from randomly selected primary care clinics (n=30), diabetology/endocrinology clinics (n=7), weight loss facilities (n=25), and fitness centers (n=30) was performed. Assessments were completed on a total of 223 scales: 94 from primary care clinics, 32 from diabetology/endocrinology clinics, 39 from weight loss centers, and 58 from fitness centers. Scales were assessed for condition, location in facility, resting surface, commercial designation, and calibration history. Scale precision was validated using 100 lb. (45.5 kg), 150 lb. (68.3 kg), 200 lb. (90.9 kg), and 250 lb. (113.6 kg) certified weights.

Results. Overall, scales demonstrated decreased precision with increased weight. At higher weights, more than 15% of scales were off by more than 6 lbs. (2.3 kg), approximately 1 Body Mass Index (BMI) unit. While facility type was not significant, condition, location in facility, resting surface, commercial designation, and calibration history were significant.

Conclusions. This study demonstrates that many scales used to measure body weight are imprecise and that scales in health care settings are no more precise than those in other facilities. Clinical decisions based on scales that are imprecise have the potential to cause iatrogenic complications in patient care.

Address correspondence to: Risa J. Stein, PhD, Department of Psychology, Rockhurst Univ., 1100 Rockhurst Rd., Kansas City, MO 64110; tel. 816-501-4691; fax 816-501-4169; e-mail <ri>risa.stein@rockhurst.edu>.

©2005 Association of Schools of Public Health

^aDepartment of Psychology, Rockhurst University, Kansas City, MO

bThe Mid-America Heart Institute, St. Luke's Hospital, University of Missouri–Kansas City, Kansas City, MO

^{&#}x27;Fairleigh-Dickinson University, Teaneck, NJ

Average individuals determine their weight using a scale, typically at a physician's office, fitness center, or weight loss center. The types of scales found at these sites are considered to be more precise than bathroom models found in many private residences. Unfortunately, few studies have evaluated the precision of non-household scales. One study, however, that examined scales used by inpatients and outpatients at a North Carolina hospital concluded that many "were too inaccurate for optimal patient care."

Although the National Institute of Standards and Technology (NIST) has established acceptable tolerance levels for scales used for commercial purposes (e.g., diet centers offering commercial weight loss programs), no such standards have been established by this agency or the Joint Commission on Accreditation of Hospitals for scales used clinically in the care of patients. Despite the lack of scale regulation for patient care, a survey of physicians suggested that, for clinical purposes, scales should be precise to ±1 lb. per 150 lbs. (45.5 kg) of body weight. Of the 97 scales examined by Schlegel-Pratt and Heizer in a North Carolina hospital, 62% would have failed to meet the NIST criteria, and 22% would have failed to meet the physician-imposed standards.

The possibility that one or more scales in a single office may produce imprecise readings that result in the presumption that substantial changes in a patient's weight have occurred, when, in fact, the patient's weight had remained stable, emphasizes the need for scale precision. Harris and colleagues, after interviewing midwives in England, reported that the midwives frequently used different scales in their clinics depending upon which one was available at the time of measurement.1 It is not unlikely that two scales in a physician's office could each be imprecise in a different direction, thus suggesting a possible weight loss at one visit and weight gain at a subsequent visit, or vice versa. Naturally, such weight variations could potentially affect medication and nutritional decisions, perspectives on fetal development, and assessments of volume status in heart failure. Indeed, after surveying midwives and scales in obstetric clinics, Harris et al. concluded that measurements taken in these clinics may not be sufficiently precise or reliable to guide decisions pertaining to weight changes associated with preeclampsia and poor fetal growth or to assess appropriate weight gain week to week during pregnancy.¹ Given that the American College of Cardiology/American Heart Association (ACC/ AHA) Heart Failure Guidelines emphasize the serial measurement of patients' weight as a Class I indication for the serial monitoring of patients, imprecise data may result in substantial iatrogenic error in the titration of medications.3

Despite obvious negative ramifications of repeated reliance on imprecise scales, few studies have addressed the precision of scales across a variety of settings. The purpose of this study was to assess the precision of scales used in commercial weight loss programs, fitness centers, and physicians' offices using governmental standards for calibration. In addition, we assessed the impact of environmental factors, such as the scales' resting surface, which also may affect precision.

METHODS

Sampling and participants

A directory of all Kansas City-area primary care physicians was obtained from a list compiled by the Missouri Department of the Healing Arts. After the approximately 15,000 listings were arranged by address to target the Kansas City metropolitan area, 30 sites were randomly selected using a random numbers table. Due to their emphasis on weight, a listing of all diabetology and endocrinology clinics in the Kansas City metropolitan area was developed based upon consultation with the Missouri Department of Healing Arts. Because fewer than 30 of these specialists were located in the region, all were included. Fitness and weight loss centers were identified through the Kansas City Yellow Pages and cross-checked with the on-line Yellow Pages before 30 of each were randomly selected using a random numbers table. The resulting pool of sites included all 30 primary care physicians' clinics, 17 endocrinology and diabetology clinics (two sites [11%] refused to participate), 25 weight loss centers (five centers [17%] refused to participate), and all 30 fitness centers.

Procedures

Four 50-lb. (22.7 kg) and two 25-lb. (11.4 kg) NIST Class F tolerance weights were purchased from a local supplier (Precise Scales Co., Inc.). Each weight was adjusted and certified to Class F tolerance (according to Military Standard [MilStd] 45662A requirements: uncertainty coverage factor k=2, 0.28 g for 50-lb. [22.7 kg] weights and 0.19 g for 25-lb. [11.4 kg] weights) by a state agency (Kansas Department of Agriculture, Division of Weights and Measures—Metrology Laboratory, Topeka, Kansas).⁴

Scale assessment training was provided by the Kansas Metrology Laboratory and required for all six scale assessment personnel. Training included methods involved in obtaining a precise reading and stacking weights in a manner representative of human weight disbursement The precision of all scales located at each site was assessed, resulting in 223 total scale precision tests (94 at primary care clinics, 32 at diabetology/endocrinology clinics, 39 at weight loss centers, and 58 at fitness centers). Factors related to scale precision were also assessed. These included scale location (male or female locker room, lobby area or public area, private area or exam room), type of scale resting surface (i.e., tile, carpet, concrete, other), overall condition of the scale (perfect condition, slightly worn, heavily worn), type of scale (digital, balance beam, other), increments of scale measurement ($\frac{1}{8}$ lb., $\frac{1}{4}$ lb., $\frac{1}{2}$ lb.), maximum capacity, commercial purpose designation, and calibration history (no history of calibration, calibrated within the past year, calibrated within the past month). Judgment of scale condition was determined by the consensus of the three raters conducting the assessment and was based upon wear and tear on the foot pad area, wearing off of numbers on the scale face or beam, and any other physical blemishes. Determining extent of wear and tear was included in the training.

Prior to placement of weights, raters assessed whether the scale was level at zero weight, and, if not, the direction in which it was off, and, in cases of over-estimation, the amount of weight it was off. After assessment at zero weight, the weight of one rater was obtained. Then, weights were placed on the scale in ascending order of 100.0 lbs. (45.5 kg), 150.0 lbs. (68.3 kg), 200.0 lbs. (90.9 kg), and 250.0 lbs. (113.6 kg). Subsequent to these assessments, the weight of the rater was again obtained.

Approach to statistical analysis

Differences between the standardized test weights and the measurement of the community scales were examined using the absolute difference between the weight of the standardized weights and the scale reading. Thus, given a standardized test weight of 100 lbs., a scale reading of 105 lbs. and 95 lbs. would have the same imprecision of 5 lbs. Absolute differences were used to prevent the canceling out of scale error in the event of over- and under-estimations. For a given location or scale characteristic, the Mean Absolute Weight Difference (MAWD) reflects the mean of the absolute differences between the weight of the standardized test weights and the scale readings. General linear modeling was used to test the difference between MAWDs among the scale sites and scale characteristics. Chi-square analyses were used to examine the relationships between scale characteristics.

RESULTS

Overall precision and precision by site

Table 1 presents weight scale imprecision, overall and by type of site, for each weight level. No significant differences were found among the various sites on measurement imprecision for any weight level. However, a repeated measures general linear model demonstrated a significant decrease in precision with increasing weight level (F-value for repeated assessment=75.38; p<0.001). For instance, the mean imprecision in weight measurement was nearly 1.9 times greater

for the 250.0 lb. (113.6 kg) test in comparison with the 100.0 lb. (45.5 kg) test (3.3 lbs. [1.5 kg] vs. 1.8 lbs. [0.8 kg]). Relatively large inaccuracies were found at all weight levels and at all locations in the sites. For example, at one primary care office using the 250.0 lb. (113.6 kg) test, a 17.5 lb. (8.0 kg) imprecision was noted. Scales found in women's locker rooms provided the only consistent under-estimations of weight.

Degree of imprecision by weight tested

Although a majority of assessed scales were precise to within 2.0 lbs. (0.9 kg) of the actual weight assessed, a significant proportion of the scales were highly imprecise, particularly at higher weight levels (see Table 2). For example, for the 250.0 lbs. (113.6 kg) test, nearly 21% of the scales were off by more than 6.0 lbs. (2.7 kg), or approximately 1 Body Mass Index (BMI, kg/m²) unit. Even for the 100.0 lbs. (45.5 kg) test, more than 25% of measurements were more than 2.0 lbs. (0.9 kg) imprecise.

Scale characteristics and precision

Table 3 presents characteristics of the scales as they relate to measurement precision. Because the 250.0 lbs. (113.6 kg) test produced the most imprecise readings and is likely the most relevant for an obese population, scale characteristics were examined at this weight level. The condition of the weight scale was related to scale precision (p<0.001), and it was found that average inaccuracies were 3.3 times greater for heavily worn scales than for scales in perfect condition. Scale precision was significantly related to scale location (p=0.014). Scales located in men's locker rooms (typically in fitness centers) were more inexact than scales in other locations. The resting surface of the scale was significantly related to the precision of the scale (p<0.001). Tile surfaces produced the greatest imprecision, while carpeted surfaces produced the least measurement error, on average. A chi-

Table 1. Absolute value of imprecision for 223 scales from 94 sites

	100.0 lbs.		150.0 lbs.		200.0 lbs.		250.0 lbs.					
	(45.5 kg) test		(68.3 kg) test		(90.9 kg) test		(113.6 kg) test					
Sites	Mean	SD	Max	Mean	SD	Max	Mean	SD	Max	Mean	SD	Max
	Ibs.	Ibs.	lbs.	lbs.	Ibs.	Ibs.	Ibs.	Ibs.	lbs.	Ibs.	Ibs.	lbs.
	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)	(kg)
All sites	1.67	2.06	10.50	1.91	2.34	12.30	2.61	3.19	14.8	3.33	4.04	17.50
	(0.79)	(0.93)	(4.73)	(0.86)	(1.05)	(5.54)	(1.17)	(1.44)	(6.66)	(1.50)	(1.82)	(7.88)
Primary care clinics	1.69	1.96	8.40	1.76	2.10	9.50	2.70	3.00	12.00	3.54	4.15	17.50
	(0.76)	(0.88)	(3.78)	(0.79)	(0.95)	(4.28)	(1.22)	(1.35)	(5.40)	(1.22)	(1.35)	(7.88)
Diabetology/	1.35	1.25	5.50	1.97	2.06	8.00	2.51	2.90	12.00	3.25	3.61	14.00
endocrinology clinics	(0.61)	(0.56)	(2.48)	(0.89)	(0.93)	(3.60)	(1.13)	(1.31)	(5.40)	(1.13)	(1.31)	(6.30)
Weight loss centers	1.33	2.38	10.50	1.35	2.27	8.00	1.97	3.59	13.50	2.23	4.15	15.30
	(0.59)	(1.07)	(4.73)	(0.61)	(1.02)	(3.60)	(0.89)	(1.62)	(6.08)	(0.89)	(1.80)	(6.89)
Fitness centers	2.04	2.32	10.50	2.40	2.70	12.30	2.98	3.28	14.80	3.79	4.06	16.50
	(0.92)	(1.04)	(4.73)	(1.08)	(1.22)	(5.54)	(1.34)	(1.48)	(6.66)	(1.34)	(1.49)	(7.43)

SD = standard deviation

Max = maximum error for any single scale

Table 2. Percentage of scales (N=223) demonstrating various degrees of
absolute weight imprecision according to weight tested

Weight tested	0.0–2.0 lbs. (0–0.90 kg) imprecision	2.1–4.0 lbs. (0.91–1.80 kg) imprecision	4.1–6.0 lbs. (1.81–2.2 kg) imprecision	>6.1 lbs. (>2.3 kg) imprecision
100.0 lbs. (45.5 kg)	74.5%	12.7%	7.7%	5.0%
150.0 lbs. (68.3 kg)	71.7%	12.6%	7.9%	7.9%
200.0 lbs. (90.9 kg)	63.0%	14.6%	7.3%	15.1%
250.0 lbs. (113.6 kg)	55.7%	16.7%	6.8%	20.8%

NOTE: Percentages may not add to 100 due to rounding.

square analysis revealed that while approximately 45% of scales found on tile surfaces were slightly or heavily worn, less than 25% of those found on carpeted surfaces were similarly worn (p=0.038). Moreover, scales were far more likely to rest on carpeted surfaces if found in private rooms (40%) or lobby areas (51%) than if placed in women's (10%) or men's (9%) locker rooms (p<0.000).

Not surprisingly, calibration of the scale in the past year resulted in scales with significantly more precise readings (p<0.001). Chi-square analysis revealed that more than twice as many scales resting on carpeted surfaces had been calibrated within the past year as compared with scales found on tile surfaces (38% vs. 15%; p=0.047). Finally, centers with a "pay by weight" distinction (always weight loss centers) produced readings that were significantly less inexact than other scales (p < 0.001).

DISCUSSION

This study demonstrates that many scales used to measure body weight provide imprecise readings. Scales in health care settings (i.e., primary care or diabetologist/endocrinologist offices) were no more precise than those in other locations (i.e., fitness or weight loss centers). Although health care standards suggest that scales not be more than 1 lb. off per 150 lbs. (45.5 kg), the mean imprecision at primary care offices for this weight was nearly 1.8 lbs. (0.8 kg) with discrepancies as great as 17.5 lbs. (8.0 kg).² Scale imprecision was strongly related to weight level, with average inaccuracies ranging from 1.3 lbs. (0.6 kg) for the 100.0 lbs. (45.5 kg) test to 3.8 lbs. (1.3 kg) for the 250.0 lbs. (113.6 kg) test.

In addition, as noted in Table 3, readily identifiable factors were associated with scale precision. For example, as a reviewer's subjective assessment of wear and tear on the scales increased, precision significantly decreased. Location also was significantly related to scale precision, with scales found in men's locker rooms being less precise than those found in women's locker rooms, lobbies, and private rooms. The resting surface of the scale also was associated with measurement precision, with scales placed on carpeted surfaces providing more exact readings than those found on tiled surfaces. Moreover, scales that had been calibrated within the past year were found to be significantly more precise than those that had not. Finally, scales designated as part of a "pay per weight" program were significantly more precise that those without this designation.

While most scales did not suffer from extreme levels of imprecision, many were imprecise at higher weight levels. For example, at the 200 lb. (90.9 kg) and 250 lb. (113.6 kg) test weights, 15.1% and 20.8% of scales were imprecise by more than 6 lbs. (2.3 kg), or approximately 1 BMI unit. Because BMI is an important metric for estimating body fat and obesity and because physicians are being encouraged to use it as a routine part of their patient assessments, weight measurements (a key component of BMI) must be precise.⁵ The greater levels of imprecision for the higher weight levels are problematic because they reduce the precision of

Table 3. Factors related to scale precision (N=223) based on 250.0 lbs. (113.6 kg) test

Factor	n	MAWD lbs. (kg)	F-value, p-value
Condition of scale			11.94, <0.001
Heavily worn	9	8.4 (3.9) ^a	
Slightly worn	62	4.2 (1.9) ^b	
Perfect condition	121	2.5 (1.1)°	
Location			3.62, =0.014
Men's locker	11	6.7 (3.0) ^a	
Women's locker	10	2.7 (1.2) ^b	
Lobby	82	2.69 (5.92)b	
Private room	89	3.57 (7.85) ^b	
Resting surface			13.94, <0.001
Tile	88	4.86 (10.69) ^a	
Carpet	92	1.88 (4.14) ^b	
Concrete	10	3.73 (8.21)	
Calibration in past year			12.05, < 0.001
No .	139	4.16 (9.15) ^a	
Yes	47	2.75 (6.05)b	
Don't Know	6	2.43 (5.35)	
"Pay by weight" designat		13.36, < 0.001	
No	169	3.71 (8.16) ^a	
Yes	22	0.46 (1.01) ^b	

NOTES: The total number of scales for each factor may be less than 223 given missing data or data falling into categories too small to analyze. Significant differences (p<0.05) occurred between superscripts (a and b; b and c; a and c) within each factor designation.

MAWD = Mean Absolute Weight Difference in pounds and kilograms.

BMI and obesity assessment in physicians' offices and weight loss/fitness centers. Thus, although these data were collected from outpatient offices, similar findings in inpatient settings could result in an increased risk of over- or undermedicating.

While the reproducibility of individual scales over time was not assessed, it seems unlikely that re-using the same scale over a short period of follow-up (i.e., one month or two) would result in substantially different results. However, different scales in the same office or health system clearly can differ substantially. Physicians who are unaware of this source of variability in weight assessment would likely assume that a patients' weight had changed significantly. If physicians were to assume, as in the case of heart failure, that the weight change was due to acute changes in volume status, diuretics might be inappropriately adjusted, potentially resulting in patient harm. Similar scenarios are possible in a range of settings and situations including obstetrics, pediatrics, and renal failure—all because patients were weighed on different scales on different occasions and physicians did not take into account this variability.

A potential limitation of this study is that it was conducted in one medium-sized metropolitan area. However, all sites within this area were randomly selected from welldefined pools making any inaccuracies documented in this study unlikely to be attributable to sampling bias. Whether Kansas City is representative of other regions, however, is unknown; therefore, similar assessments should be replicated elsewhere. In addition, the overwhelming majority of scales included in this study were of the balance-beam type. Larger studies should assess whether this type of scale is any more or less precise than other available types of scales. Furthermore, scales used in inpatient and pediatric populations should be assessed for precision. Finally, future studies should examine the temporal stability of body weight scale measurements to determine whether community scales tend to become more precise or imprecise over time.

Given the apparent simplicity of measuring weight and its importance in clinical practice, physicians occasionally may find themselves placing too much faith in a measure that may be inexact. To overcome this pitfall, physicians should ensure that their scales are calibrated regularly, that they rest upon a carpeted surface, if possible, and that they are in good apparent condition. Moreover, office personnel should be advised to make every attempt to weigh patients on the same scale at each visit. This could easily be accomplished by designating one scale for patients whose names fall in the first half of the alphabet and another for those with names falling in the bottom half of the alphabet. Following these simple guidelines could potentially increase the precision of weight measurement and enhance the quality of patient care.

REFERENCES

- Harris HE, Ellison GT, Holliday M, Nickson C. How accurate are antenatal weight measurements? A survey of hospital and community clinics in a South Thames Region NHS Trust. Paediatr Perinat Epidemiol 1998;12:163-75.
- Schlegel-Pratt K, Heizer WD. The accuracy of scales used to weigh patients. Nutr Clin Pract 1990;5:254-7.
- Hunt SA, Baker DW, Chin MH, Cinquegrani MP, Feldman AM, Francis GS, et al. Evaluation and management of heart failure. ACC/AHA guidelines for the evaluation and management of chronic heart failure in the adult [cited 2004 Dec 28]. Available from: URL: http://www.acc.org/clinical/guidelines/failure/hf index.htm
- National Institute of Standards and Technology, Weights and Measures Division. NIST handbook 145, SOP No.8 [cited 2005 Jan 20]. Available from: URL: http://ts.nist.gov/ts/htdocs/230/235/weight cart/CartBackground.htm
- National Institutes of Health/National Heart, Lung, and Blood Institute (NIH/NHLBI). Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report. Washington: U.S. Government Press; 1998.